Available online at www.sciencedirect.com

ScienceDirect

Vision Research 46 (2006) 4321–4332

Vision Research

www.elsevier.com/locate/visres

Disparity with respect to a local reference plane as a dominant cue for stereoscopic depth relief

Yury Petrov [a,*], Andrew Glennerster [b,c]

[a] *Psychology Department, 125 Nightingale Hall, 360 Huntington Ave., Northeastern University, Boston, MA 02115, USA*
[b] *School of Psychology and Clinical Language Sciences, University of Reading, Reading RG6 6AL, UK*
[c] *Department of Physiology, Anatomy and Genetics, Sherrington Building, Parks Road, Oxford OX1 3PT, UK*

Received 5 January 2006; received in revised form 15 July 2006

Abstract

Earlier studies showed that the disparity with respect to other visible points could not explain stereoacuity performance, nor could various spatial derivatives of disparity [Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence of surface-based processing of binocular disparity. *Current Biology*, 12:825–828; Petrov, Y., & Glennerster, A. (2004). The role of the local reference in stereoscopic detection of depth relief. *Vision Research*, 44:367–376.] Two possible cues remain: (i) local changes in disparity gradient or (ii) disparity with respect to an interpolated line drawn through the reference points. Here, we aimed to distinguish between these two cues. Subjects judged, in a two AFC paradigm, whether a target dot was in front of a plane defined by three reference dots or, in other experiments, in front of a line defined by two reference dots. We tested different slants of the reference line or plane and different locations of the target relative to the reference points. For slanted reference lines or plane, stereoacuity changed little as the target position was varied. For judgments relative to a frontoparallel reference line, stereoacuity did vary with target position, but less than would be predicted by disparity gradient change. This provides evidence that disparity with respect to the reference plane is an important cue. We discuss the potential advantages of this measure in generating a representation of surface relief that is invariant to viewpoint transformations.

Keywords: Psychophysics; Stereopsis; Disparity; Curvature

1. Introduction

Studies of stereoacuity performance have been used as an effective probe to investigate the processing and representation of disparity information in the visual system. Despite some important advances in our understanding, however, there is currently no agreement about the stereoscopic cue (or cues) that the visual system uses to detect depth relief.

In an early study on the subject (Westheimer & McKee, 1979), relative disparity was suggested as the principal cue in stereoscopic depth discrimination. The stimulus was a simple three-line configuration comprising a target line positioned in between two reference lines displayed at the same depth. In a later work stereoacuity was shown to decrease with separation between the target and the reference line (McKee, Welch, Taylor, & Bowne, 1990). Although the effect was explained by eccentricity considerations, disparity gradient as a possible cue also predicts this behavior. Disparity curvature (second spatial derivative of disparity) (Rogers & Cagenello, 1989) and even more complex differential cues (Lappin & Craft, 2000) were suggested as alternative disparity cues limiting stereoacuity.

Other studies suggest that the stereo system uses some measure of disparity relative to a locally defined reference frame at an early stage in processing. For example, in Glennerster, McKee, & Birch (2002), a regular grid of dots slanted around the vertical axis was presented, and the subjects' sensitivity for detecting displacements of the central column of dots was measured. The results demonstrate that

* Corresponding author.
E-mail address: yury@ski.org (Y. Petrov).

0042-6989/$ - see front matter
doi:10.1016/j.visres.2006.07.011

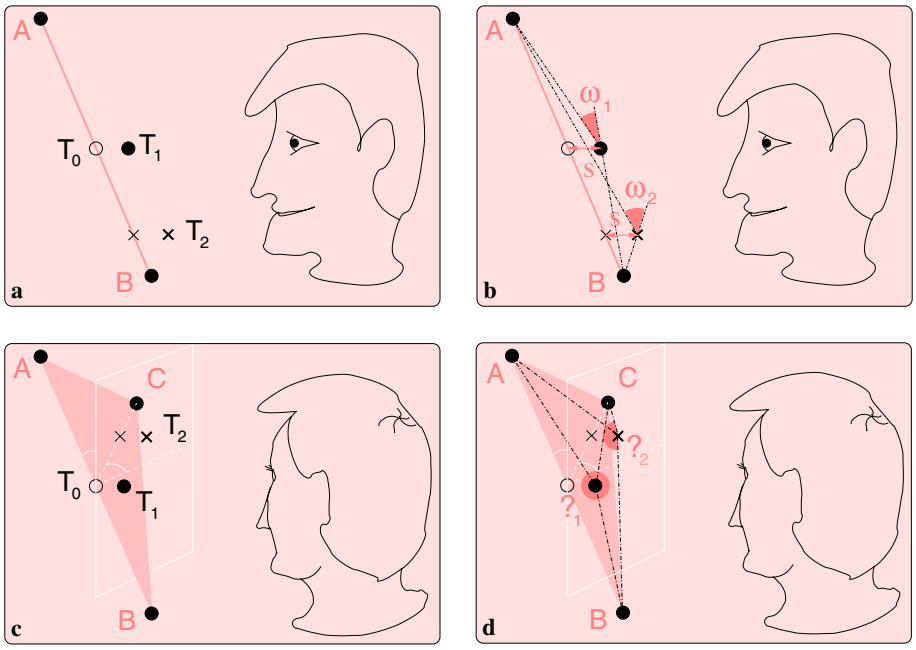

Fig. 1. A schematic diagram of the stimuli used in this study and potential disparity cues. (a) The triangle ABT defines a two-dimensional depth profile (side view). The observer's task was to identify the interval in which the target was displaced from the line AB, e.g. T_1 versus T_0 or, in a different trial, T_2 versus the cross shown on line AB. (b) Disparity cues the observer might use to carry out the depth discrimination shown in (a). These include disparity s with respect to the reference line AB, and disparity gradient difference ω (both category IV cues, see text). The disparity gradient difference at position 2 (ω_2) is larger than that at position 1 (ω_1), while disparity s remains the same. (c) The tetrahedron $ABCT$ defines a three-dimensional depth 'bump'. The plane ABC is shown in gray. The thin white line shows the fronto-parallel plane through T_0. As in (a), T_0 is a target location on the line AB while T_1 has a disparity s with respect to it. Likewise, T_2 has a disparity s with respect to the cross marked on the line T_0C. (d) Disparity cues that could be used in (c). Two sets of three disparity gradients (AT, BT and CT) describing the surface depth relief for target positions T_1 and T_2. '?' denotes some (unknown) 3D measure based on the three disparity gradients analogous to ω.

the sensitivity was determined by the distance of the column from the plane of the grid or some similar cue, rather than by the change of its relative disparity.

In our previous work (Petrov & Glennerster, 2004) we used this kind of stimulus in its minimal form: two reference dots defined a reference line that was slanted in depth and a target dot whose position had to be judged in relation to that line. Fig. 1a illustrates the stimulus. Two reference points, A and B, define a slanted line. The subject's task was to detect when the target dot T had been displaced from location T_0 on the line, to the test location T_1 in front of the line. We were interested in the cue that the visual system uses to detect this displacement.

Previously, we classified the possible cues in our task into four categories based on the order of spatial derivatives of the disparity signal:

- I: zeroth order, e.g. relative disparities between the target dot T and reference dots A and B.
- II: first order, e.g. disparity gradients between the target dot T and reference dots A and B.
- III: second order, e.g. change in disparity gradient at T normalized by the angular distance AB (which is a measure of disparity curvature).
- IV: other disparity cues: e.g. (a) disparity gradient difference at the target point T and (b) disparity relative to the reference line AB.

The previous study demonstrated that only cues from the last category define performance in our task.[1] The two cues can be separated by changing the position of the target along the line AB. This point is illustrated in Fig. 1b, where the target in position 2 (T_2) is shown to create a larger disparity gradient difference ω as compared to the target in position 1 (T_1). At the same time, disparity s relative to the AB line is the same for both positions. The disparity gradient difference is illustrated in Fig. 1b by the angle ω formed by imaginary lines connecting the target dot T with reference dots A and B. (In fact, for small disparities s used in the present study, the disparity gradient difference is essentially equal to ω measured in radians, so we will use them interchangeably.) The Appendix A gives a derivation of the relation between ω and s for different positions of the target. This method was used in experiments 1 and 2 to determine whether the target positioning along the AB line has any effect on stereoacuity.

Of course, the more general case is the detection of a surface 'bump' rather than relief with respect to a line (2D depth profile). Accordingly, we tested whether the position of the target on a general 3D plane defined by

[1] Note that for each of the cues the underlying measures used by the visual system could be related to them in an arbitrary monotonic way.

three reference dots affected stereoacuity. However, in this case some of the cues described above do not have a straightforward generalization. For example, the change in disparity gradient for the case of a 2D depth profile shown in Fig. 1b, becomes ill-defined. Three reference points A, B, and C shown in Fig. 1c together with the target T form a surface bump in depth. There are three disparity gradients (AT, BT, and CT) as well as three angles between the directions of the gradients now, and it is not clear how to form a meaningful disparity measure using these quantities (this point is illustrated by question marks in Fig. 1d). Nevertheless, any such measure would grow fast as the target approaches one of the reference dots.

Our results show that, in most cases performance is affected remarkably little by the target's position with respect to the reference dots. This pattern is consistent with the disparity relative to the reference line or plane (s) dominating performance in the general case.

2. Method

2.1. Apparatus

Stimuli were generated on a Sun Ultra-10 Workstation and displayed on two high-resolution color monitors (Flexscan T961, Eizo). Stereo images were viewed via a modified Wheatstone stereoscope at a viewing distance of 2.65 m. The display was 1600×1280 pixels, and each pixel subtended 18 s of arc. Anti-aliasing of circular dot edges was used to generate sub-pixel resolution. Stimuli were viewed in a dark room. The background luminance was very low (<1 cd/m^2), and the stimuli were bright (60 cd/m^2).

2.2. Subjects

All three observers had normal or corrected monocular visual acuity and were experienced stereo-observers (AF was naive to the hypothesis tested). The subjects were allowed to train for 30 min before the beginning of the experiment.

2.3. Psychometric procedure

We used a two-interval forced-choice paradigm. A stimulus, such as that shown in Fig. 2b, was displayed in one interval (I) with the target dot T placed on the line defined by two reference dots A and B (position 0, shown by the open circle T_0 in Fig. 2a). In the other interval (II), the target dot was displaced toward the observer (position 1, shown by the filled circle T_1 in Fig. 2a). The presentation order was randomized and the subject's task was to identify the interval in which the target dot had been displaced. For each subject the displacement magnitude was chosen approximately at his/her detection threshold, i.e. at 75% correct performance, as measured in a pilot experiment. The magnitude of the target disparity was 14″, 18″, and 18″ for subjects YP, AF, and AG, respectively. Although the displacement was zero in one of the intervals, the stimulus normally appeared to have a small degree of depth curvature in both intervals, sometimes even in the direction opposite to the one actually shown. This is characteristic of sub-threshold and near-threshold perception. Subjects were instructed to indicate the interval in which the target dot appeared closer.

Each of the two stimulus intervals lasted 1.5 s, during which the subject was free to move their eyes. The screen was blank in the 1 s inter-stimulus interval. Before the first stimulus interval, a fixation stimulus was presented for 1 s. It consisted of a central diamond outline (36 arcmin) and four bright dots forming a 4° square, also centered on the midpoint of the screen. This stimulus provided a visual reference for fronto-parallel. After the second interval the screen was blank until the subject gave their response which triggered the next stimulus to be displayed.

In a run of 100 trials, five different target positions along a reference line (experiments 1, 2, and 3) and a reference plane (experiments 4 and 5) were tested (marked by crosses in Fig. 2a and c), with the trials testing each position randomly interleaved. The data in Figs. 3–7 show the proportion of correct responses made over at least 200 trials for each condition. Error bars in these plots show the standard error of the mean computed from the binomial distribution.

2.4. Stimuli

In all of the experiments, the separation between the reference dots A and B was 86 arcmin as viewed from the cyclopean point. The relative disparity between them was 8.6 arcmin, i.e. the disparity gradient was 0.1. The dots were 2.5 arcmin in diameter, blurred with a Gaussian kernel for the purposes of anti-aliasing. In experiments 4 and 5 a third reference dot C was added to define a reference plane ABC, as shown in Fig. 2c. The position of the third reference dot was chosen so that the three dots formed an equilateral triangle as viewed from the cyclopean point. The cyclopean separation between C and the middle point of the AB line (T_0 in Fig. 2c) was 74.5 arcmin, while the relative disparity was 7.45 arcmin. Thus, the reference plane was characterized by a disparity gradient of 0.1 both along the vertical and the horizontal axis, and appeared slanted in depth (Fig. 2c and d).

In experiments 1, 2 and 3 (two reference dots) the target dot T was displayed in five positions located at equal intervals along the reference line AB. The five target positions marked by the open circle and the thin crosses in Fig. 2a correspond to interval I, in which the target was placed on the reference line. The closed circle and the fat crosses indicate the corresponding target position in interval II, when the target was displaced from the reference line toward the observer. Fig. 2b illustrates three of the ten possible stimulus configurations: (0) target at the middle point of the reference line (interval I), (1) the same but for interval II, (2) target at the bottom position (interval II).

In experiments 4 and 5 (three reference dots) the target dot was displayed in five positions located at equal intervals along the reference plane ABC (Fig. 2c). Again, the thin crosses indicate target positions in interval I, and the fat crosses the corresponding positions in interval II. In this case the target locations were aligned horizontally along the line T_0C (Fig. 2c). Fig. 2d illustrates three of the ten possible stimulus configurations: (0) target at the leftmost position (the middle point of the AB line, interval I), (1) the same but for interval II, (2) target at the rightmost position (interval II).

3. Results and discussion

The experimental results are shown in Figs. 3–7. Detectability[2], d', was plotted on the y-axis versus the target position relative to the reference dots plotted on the x-axis. The first three subplots present the results for each subject separately, the last subplot shows the average performance.

In all the experiments, we examined whether the location of the target with respect to the reference dots affected stereoacuity performance, since this manipulation gives rise to different predictions according to the two hypotheses described in the Introduction. These are that the principal cue used by the visual system is (i) disparity with respect to a reference line or plane or (ii) disparity gradient difference

[2] Detectability, d', was defined by $d' = \sqrt{2}F^{-1}(P)$, where P was the proportion of correct answers (when target displacement was detected), and F^{-1} was the inverse of the cumulative Gaussian function.

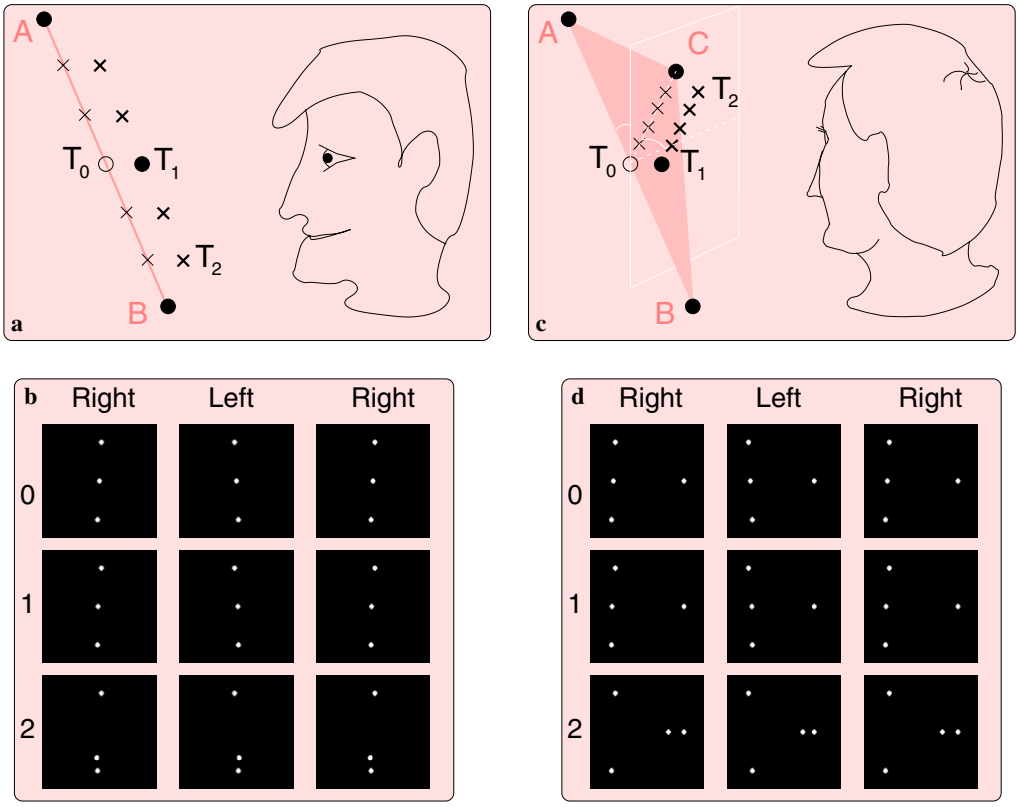

Fig. 2. (a) As Fig. 1a, but showing the five different locations along the reference line *AB* that were tested within one run. (b) Stereo pairs of the stimulus. The left and right eye images in the first two columns are arranged for crossed fusion (for uncrossed fusion use the last two columns). Stimuli with the central dot in positions 0, 1, and 2 are shown in the first, second and third row, respectively. (c) A schematic diagram of the four-dot stimulus used in experiments 4 and 5 (see Fig. 1c). Again, five dot locations along the *ABC* plane are shown. These were tested on randomly chosen trials within one run. (d) Stereopairs of the stimulus with the central dot in positions 0, 1, and 2 are shown in the first, second and third row, respectively. The slants shown schematically in (a) and (b) correspond approximately to the perceived slants of the stimulus (which are significantly smaller than the metric slants for the viewing distance in the experiment).

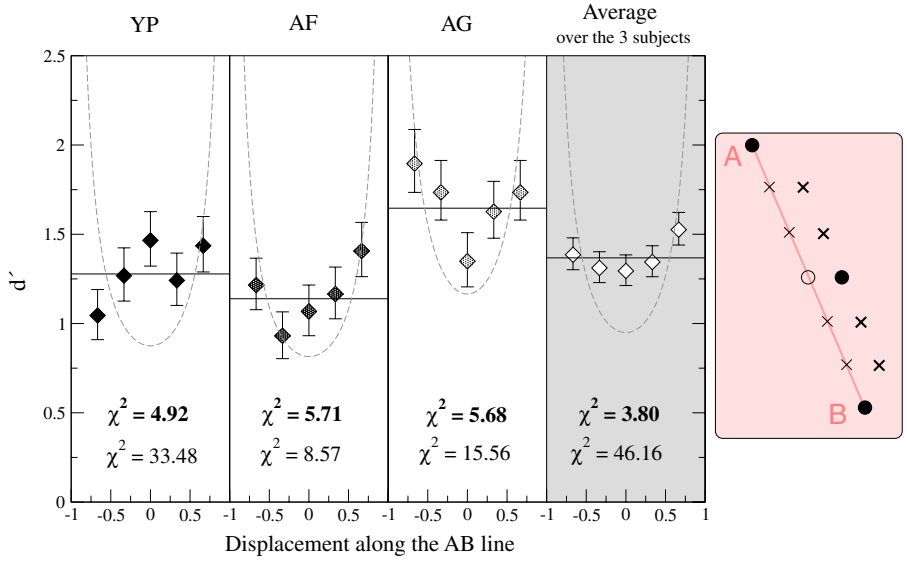

Fig. 3. Results of experiment 1. The reference line *AB* was slanted in depth (shown in side-view in the rightmost panel). Detectability d' of a depth profile with respect to a reference line is shown on the y-axis as a function of the target position along the reference line *AB* (*A*'s position was defined as 1, *B*'s as −1). The first three subplots show results for three observers, the fourth subplot shows the averaged data. Constant d' fit is shown with black lines (with χ^2 values shown in bold), the disparity gradient difference (ω) fit is shown with dashed gray curves (with χ^2 values shown in plain text). Error bars in all figures show standard error of the mean.

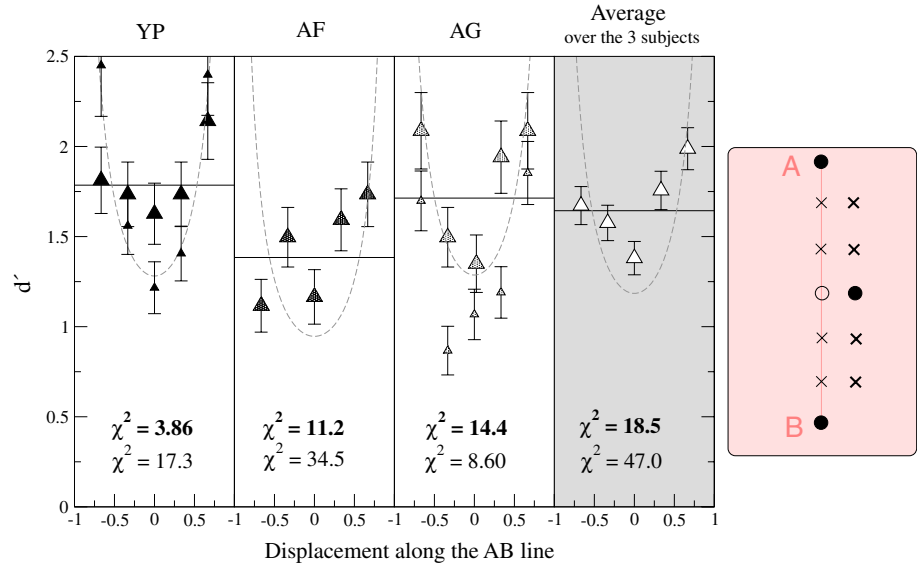

Fig. 4. Results of experiment 2. The reference line AB was fronto-parallel. Small symbols for observers YP and AG show the results of a control experiment with only one reference dot (either A or B, depending on which one was closest to the target). The constant d' fit is shown with black lines, disparity gradient difference (ω) fit is shown with gray curves, as in Fig. 3.

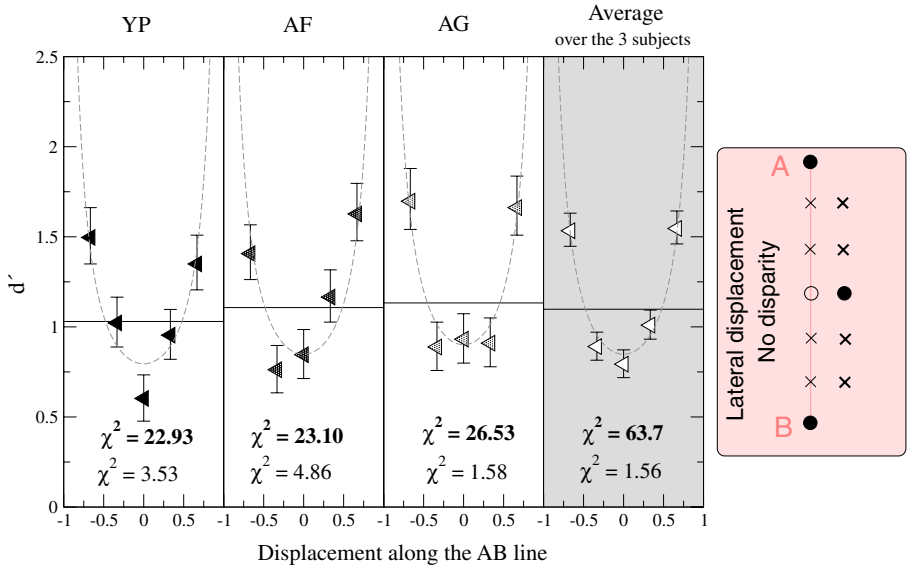

Fig. 5. Results of experiment 3. The same stimulus as in experiment 2, but the left-eye image was shown to both eyes. Monocular target displacement was twice as large as in experiment 2. Again, the constant fit is shown with black lines, ω fit with gray curves.

at the target. Fits corresponding to the predictions of hypothesis (i) are shown in Figs. 3–7. A constant d' is predicted by this hypothesis, because in the signal interval the target disparity with respect to the reference line or plane was constant for all target locations. Hypothesis (ii) predicts, on the other hand, that performance should vary with target location according to the disparity gradient difference ω (defined for a depth profile with respect to a reference line), or some analogous surface measure (see Introduction). In particular, this second hypothesis predicts that performance should be greatly improved as the target is moved closer to a reference point. Fits based on ω cue are shown in Figs. 3–5.

Both for constant d' fit and ω fit the respective functional form was multiplied by a free parameter, and the least-squares goodness-of-fit measure χ^2 was minimized to find the optimal value of the parameter. Solid horizontal lines show the predictions of the constant fit (hypothesis (i)); dashed gray curves show the disparity gradient prediction (hypothesis (ii)). The corresponding values of χ^2 are shown in bold and plain text, respectively, at the bottom of each plot.

3.1. Experiment 1: Slanted reference line

We began by testing stereoacuity at different positions along a slanted reference line (line AB in Figs. 1a and 2a).

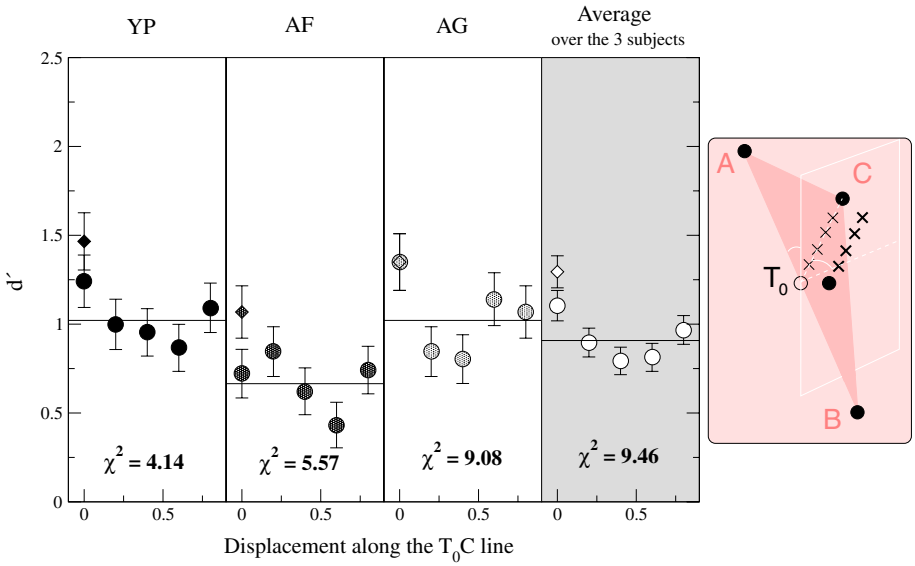

Fig. 6. Results of experiment 4. The detectability d' of a surface depth 'bump' (illustrated in the rightmost panel) is shown on the y-axis as a function of the target position along the line T_0C which lies in reference plane ABC. The middle point of the line AB was defined as 0, the C dot position was defined as 1. Diamonds show a corresponding $x = 0$ datum from Fig. 3. Constant d' fit is shown with black lines.

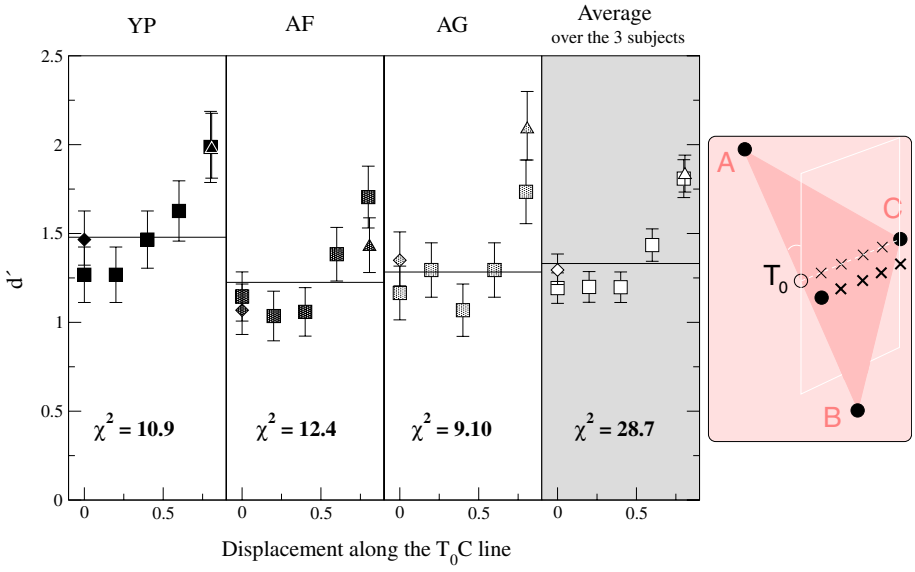

Fig. 7. Results of experiment 5. Similar to Experiment 4, except that now the reference dot C in the $ABCT$ tetrahedron was positioned at the same depth as the target T in one of the AFC intervals (as illustrated in the rightmost panel). Diamonds show a corresponding $x = 0$ datum from Fig. 3. Triangles show relevant data from Fig. 4: the average performance for the two target positions closest to A and B reference dots. Constant d' fit is shown with black lines.

The results are shown in Fig. 3; the target position is plotted along the x-axis relative to the locations of reference dots A and B: the positions of dots B and A are defined as -1 and 1, respectively. One can see that the performance did not change significantly as the position of the target was varied, although for two subjects there was a tendency for performance to improve at positions near the reference points. The data were fitted with a constant disparity fit $d' \sim s$ (where s is disparity of the target with respect to the line AB) and the disparity gradient difference fit $d' \sim \omega$ (see Appendix A for the details). For all observers the constant fit shown by a solid black line had a much better agreement with the data than the ω fit shown by a dashed curve. For the constant d' fit, the χ^2 were always well below the 95% confidence interval ($\chi^2_{0.95} = 9.48$, 1 degree of freedom). In contrast, the χ^2 values obtained from the disparity gradient difference fit were generally large. The superiority of the constant d' fit is especially clear in the plot of the average data for all three subjects.

Although the data do not rule out the possibility that performance close to the reference points is in part determined by the disparity gradient difference, it is clear that, overall, disparity relative to the reference line AB is the dominant cue in this task.

An alternative viewpoint is that disparity gradient difference is the dominant cue over the whole range of target positions but that the predicted sharp increase in performance close to the reference dot is not observed in our data because 'crowding' or lateral interactions diminish stereoacuity close to a reference dot. However, Westheimer & McKee (1979) show that stereoacuity improves (if anything) down to about 5 arcmin and then deteriorates precipitously for smaller separations while our smallest separation was 17 arcmin, well above this limit.

3.2. Experiment 2: Fronto-parallel reference line

We next explored the special case in which the reference line AB is fronto-parallel. In this case, there is a much greater chance that the relative disparity (or disparity gradient) between two points could act as a helpful cue. This is because stereo is especially sensitive when target and comparison stimulus are at very similar disparities. When target and comparison stimuli are at quite different disparities (as is the case for the slanted stimuli in experiment 1), stereoacuity thresholds for depth increment thresholds are known to be poor (Blakemore, 1970; McKee et al., 1990). This might explain why another cue, such as disparity with respect to the reference line, dominates performance in the case of the slanted stimulus. In order to test this possibility, the stimulus was modified by setting the slant of the AB line to zero. If the stereoacuity threshold is determined solely by the relative disparity between the target and the reference line AB, one would expect performance in Experiment 2 to be the same as in Experiment 1. If, on the other hand, additional cues were used, then performance should improve. If disparity gradient is used, then performance should also depend on position along the reference line, AB.

As Fig. 4 shows, performance for observers YP and AG (large symbols) was significantly higher compared to that in experiment 1. This is in agreement with earlier studies (e.g. Lappin & Craft, 2000) and indicates that other disparity cues besides disparity relative to the reference line AB affected stereoacuity. Because for AG $\chi^2_{\text{const}} > \chi^2_{\omega}$, the performance was better described by the $d' \sim \omega$ fit (shown by a gray curve) than by the constant fit. Note that in experiment 1 AG also demonstrated a slightly better performance when the target was positioned close to either of the reference dots (Fig. 3), but in the present case the increase in performance was more significant. The same tendency can be seen in YP's data in this experiment.

Type II cues, e.g. disparity gradient, are likely to produce such an effect. To check this possibility, observers YP and AG repeated the present experiment with one of the reference dots omitted (the remaining dot was always the one closest to the target). Cues based on disparity gradient difference as well as disparity with respect to the line AB were thus removed. As before, the observer's task was to report in which of the two AFC intervals the target appeared closer to the observer.

The results for the modified stimulus are shown in Fig. 4 by small triangles. One can see that performance dropped significantly, when the target was moved away from the remaining reference dot. This suggests that for the special case in which the reference line AB is fronto-parallel, type II cues can provide a significant input in addition to the relative disparity s cue when the target is close to one of the reference dots.

3.3. Experiment 3: Monocular detection

Besides stereoscopic cues, monocular lateral displacement of the target dot accompanying the change of the target disparity could provide additional cues. In this experiment, sensitivity for monocular displacement of the target dot was studied as a function of the target position relative to the reference dots A and B. The stimulus from experiment 2 was used here with the left-eye image shown to both eyes. The observer's task was to indicate in which of the two AFC intervals the target was displaced to the right of the line AB. Even though the target was always positioned on the AB line in the non-signal interval (interval I), subjectively it often appeared to be set off either way. Therefore, the observer's task was formulated in a different way: to indicate in which of the two intervals the target dot appeared rightmost. Because earlier studies showed that monocular displacement thresholds are several times higher than stereoacuity thresholds (e.g. Westheimer & McKee, 1979), the target lateral displacement was set to twice the monocular displacement used for each subject in the stereoscopic task (e.g. 14″ displacement for observer YP).

Results for the monocular task are shown in Fig. 5. In agreement with the earlier studies monocular detectability was found to be much lower than stereoscopic detectability (considering that lateral displacement was increased by a factor of two in the present experiment). This indicates that the effect of monocular cues for the stereoscopic task was negligible. Interestingly, in contrast to stereoscopic performance, detectability increased dramatically as the target was brought closer to either of the reference dots. The data were well fitted by the ω fit ($\chi^2 < \chi^2_{0.95}$ for all subjects). Previously we have defined ω as a disparity gradient but the predictions can be applied, by analogy, to the 2D case. For small target displacements in the frontoparallel plane which are considered here, ω describes deviation from straightness in the ATB line. More specifically, it gives twice the deviation of the ATB angle from 180 degrees (see Fig. 10). This demonstrates that detection of monocular displacement is quite different from its stereoscopic counterpart, and is likely to be based on differential cues (such as ω), rather than on cues based on displacement with respect to the reference line AB.

3.4. Experiment 4: Slanted reference plane

Earlier work on stereoacuity was based on stimuli containing one or two reference bars or dots (e.g. McKee et al., 1990; Westheimer & McKee, 1979) where the stimulus was essentially a 2D depth profile. In recent work by Glennerster & McKee (1999) & Glennerster et al. (2002) a square grid of dots was used as a stimulus, but because a whole column was used as a target, the stimulus was equivalent to a single row of dots.

It would be interesting to study a more general case of surface depth relief. In its minimal form it is a depth 'bump' constructed of four dots forming a tetrahedron. To this end, stimuli used in experiments 1 and 2 were modified by adding a third reference dot (C) as described in Section 2.4 and shown in Fig. 2c. The magnitude of the disparity gradient between points A and B characterizing the vertical slant of the ABC plane remained the same as in experiment 1 (0.1). In addition, the plane was slanted horizontally, with the disparity gradient between points T_0 and C (Fig. 2c) also equal to 0.1. Observers reported perceiving the stimulus as a plane slanted in depth with the target dot protruding out of the plane thus forming a shallow pyramid.

As can be seen from Fig. 6, results for this stimulus were similar to those in experiment 1, although for all observers performance deteriorated somewhat. There was no significant dependence on the target's position inside the ABC triangle. The position was measured from the middle point of the AB line (defined as 0) to the C dot position (defined as 1). The constant fit (the null hypothesis) was acceptable for all subjects ($\chi^2 < \chi^2_{0.95} = 9.48$), but the agreement was poorer than in experiment 1, mainly because performance at position 0 was on the average significantly better than for the other locations.

Because the target at position 0 falls onto the AB line, it is likely that for this particular location the AB line was used as a reference (instead of the ABC plane), which explains the difference in performance for this location. The related data point from experiment 1 is shown by a diamond in each subplot of Fig. 6 for comparison. One can see that performance at position 0 was quite close in the two experiments.

Altogether, the results of this experiment extend the results of experiment 1: given a local reference, such as a line defined by two reference dots or a plane defined by three reference dots, stereoscopic detection of a depth 'bump' is dominated by its disparity relative to the local reference, rather than by the change of disparity gradient over the bump's surface.

3.5. Experiment 5: Target positioned at the same depth as the reference dot C

Experiment 2 demonstrated that when reference dots are frontoparallel and positioned at the same depth as the target in one of the AFC intervals, the disparity gradient between the target and the closest reference dot becomes an important factor in determining stereoacuity. Here we show that the same applies for the stimulus with three reference dots. The stimulus was modified by setting the relative disparity between point C and the middle point of the AB line to zero. Thus, the horizontal component of the ABC plane's slant was zero, and the target always had the same disparity as dot C in one of the AFC intervals.

The results for this stimulus are shown in Fig. 7, where the position of the target dot plotted along the x-axis was measured in the same way as in the previous experiment. The effect of having the reference dot C at the same disparity as the target is quite obvious: performance sharply increased as the separation between the target and the reference dot C decreased. As a consequence, the constant fit was not acceptable ($\chi^2 > \chi^2_{0.95} = 9.48$) for observers YP and AF, and for the averaged data.

As in Fig. 6, performance for the target position 0 was compared with performance in experiment 1 (shown with a diamond in each subplot). Also, performance for the target position closest to dot C was compared with relevant data in experiment 2: the average performance for the two target positions closest to A and B reference dots is shown with triangles. The data from both experiments 1 and 2 are close to the results of this experiment, indicating that disparity with respect to the reference line AB and disparity gradient with respect to a reference dot C, respectively, are likely to be used as the principal cues at these locations.

4. General discussion

4.1. Main results

There are three main results of this work. First, the results reported in this paper extend previous findings showing that relative disparity, disparity gradient and disparity curvature cues (category I, II, and III described in the Section 1) cannot explain performance in a stereoscopic detection task. Instead, two other cues were suggested as primary (Petrov & Glennerster, 2004): disparity gradient difference or disparity with respect to the reference line (category IV cues). Here, we have presented experimental evidence showing that of the two cues, disparity with respect to the reference line dominates the performance in this task. The disparity gradient difference at the target position predicts a large increase in performance when the target is moved closer to either of the reference dots, which, in general, does not agree with the experimental data (Fig. 3).

Second, we have shown that performance in this task cannot be explained by monocular displacement cues. Monocular performance was substantially lower than stereoscopic performance and also varied strongly with the target position along the reference line (Fig. 5) unlike stereo performance (Figs. 3 and 4).

Third, we replaced a depth profile with respect to a line (Fig. 2b) by a more general surface depth relief in the form of a tetrahedron *ABCT* (Fig. 2d). In the general case, the data are consistent with disparity relative to the reference plane *ABC* dominating performance in this task. We found some exceptions to this general rule, both in the case of the reference line and reference plane. The exceptions are consistent with the visual system combining both types of cue, as discussed below.

4.2. Stereoscopic cue combination

Although in general disparity with respect to the reference line (or plane), *s*, appears to be a good predictor of our data, there is one case in which it does not describe it well. When a reference point was positioned at the same depth as the target (in the non-signal AFC interval), subjects performed significantly better if the target was close to this reference point (Figs. 4 and 7). This suggests that disparity gradient with respect to this point acted as an additional cue in this case.

A control experiment, in which only one reference point was left, confirmed that disparity gradient could indeed provide a sensitive cue when the target was close to the reference point and at the same depth. This is consistent with other stereoacuity studies (e.g. Rogers & Graham, 1982; Tyler, 1974).

Thus, it is likely that a combination of two cues was used: disparity with respect to the reference line (reference plane in experiment 5) and disparity gradient. Our hypothesis is that in experiment 1 and 4 the disparity gradient cue is sufficiently poor that it should add little to the predicted d', hence performance is dominated by disparity with respect to the reference line. It is reasonable to assume that the disparity gradient cue is weak when the reference line is slanted since the visual system must detect small differences in disparity gradient in the presence of a large pedestal disparity gradient. On the other hand, in experiments 2 and 5 the disparity gradient of the reference line (*AB* in experiment 2 and T_0AC in experiment 5) is zero. Hence, according to Weber's law, disparity gradient changes should be more detectable. In Fig. 8, data from experiments 2 and 5, averaged between subjects, are replotted from Figs. 4 and 7. Also, data from experiments 1 and 4 (Figs. 3 and 6) are shown for comparison in the two shaded subplots.

Fig. 8 demonstrates that observers do indeed use more than disparity with respect to the reference line or plane.

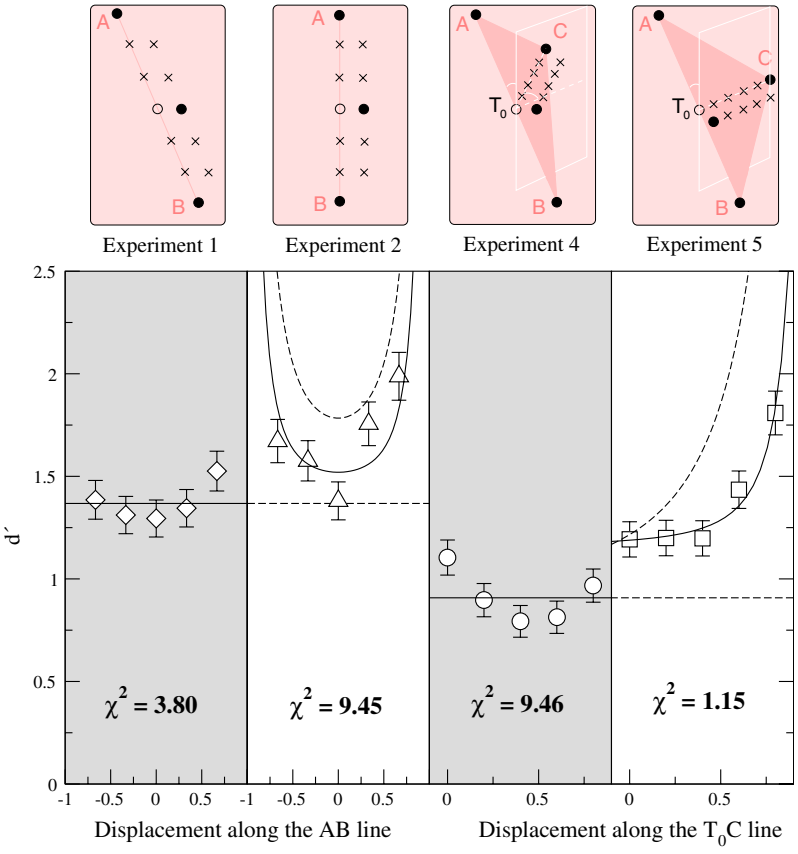

Fig. 8. Data re-plotted from Figs. 3, 4, 6, and 7 showing only the averaged data in each case. Dashed lines indicate performance in experiments 2 and 5 being limited from below by performance in experiments 1 and 4, respectively, where the disparity *s* with respect to reference line (or plane) was the only cue. Dashed curves predict performance if the disparity *s* and disparity gradient cues were optimally combined. Solid curves show the best fit to the performance in experiments 2 and 5 assuming a linear combination of disparity *s* and disparity gradient cues.

Another cue, such as disparity gradient difference, must contribute significantly to performance when the reference line is fronto-parallel and the target is close to one of the reference dots. These are the situations when the disparity gradient difference, ω, would be expected to provide a particularly reliable signal. If the two cues, s and ω, were combined optimally to maximize performance, and had statistically independent noise, total detectability would be equal to $\sqrt{d'_s{}^2 + d'_\omega{}^2}$, where d'_s is detectability due to disparity s relative to the reference line (or plane), and d'_ω is detectability due to disparity gradients. d'_s is provided by the data in experiments 1 and 4, while d'_ω was obtained using results for observers YP and AG in the control experiment described in Section 3.2. The upper limit of performance calculated in this way is shown in Fig. 8 by dashed curves. If, on the other hand, disparity s alone was used in all the four experiments, performance in experiments 2 and 5 should be equal to that in experiments 1 and 4, respectively. This lower limit is shown by dashed lines.

Clearly, performance in experiments 2 and 5, was intermediate between the two limits. It was significantly higher than the 's-only' limit, but also much poorer than predicted by the optimal cue combination. Our results were well fit by a linear combination of the two cues, where two free parameters were used as linear coefficients for d'_s and d'_ω. The choice of a linear combination was somewhat arbitrary as there are, of course, many sub-optimal ways to combine the information from two cues. The fitted curves and corresponding χ^2 values are shown in Fig. 8. The fact that the data are below that predicted by optimal combination when both depth cues are based on the disparity signal is compatible with earlier evidence (e.g. Glennerster & McKee, 1999). In contrast, when depth cues come from different features, such as disparity and texture, the combination is close to optimal (Knill & Saunders, 2003). It is possible that in the former case the suboptimal cue combination results from the two visual cues being calculated using some of the same ingredients. (Eq. (3) in the Appendix A emphasises the close relationship between the two measures.) In this case the two cues would be likely to have at least some correlated sources of noise, and the d' summation used to calculate the upper limit in Fig. 8 for uncorrelated noise would overestimate the actual performance.

4.3. Benefits of disparity measured with respect to a surface-centered reference frame

Earlier works in which stereoacuity was used to probe the underlying mechanism of depth encoding include Lappin and Craft (2000) and Rogers and Cagenello (1989). In Lapin and Craft's study, stereoacuity was tested under a variety of stimulus transformations, including the overall stimulus jitter and slant in depth. Because stereoacuity remained largely unaffected, it was concluded that a rather complicated shape measure involving joint second-order space differentials in two directions was needed to explain the results. Along with Rogers and Cagenello (1989) this suggests that the visual system adopted a strategy, according to which shape is at some (early) stage encoded by a set of local spatial differentials (e.g. disparity curvature). In fact, disparity with respect to a reference line or plane suggested here conforms with these experimental results equally well.

We have presented our data as evidence that 'differential' cues alone (i.e. disparity gradient difference ω or analogous cues) cannot explain the results. Instead, disparity with respect to the reference line or the reference plane were suggested as the primary cues. One important benefit of such 'referential' shape description is that it represents the underlying surface relief in a more straightforward way. Firstly, the reference plane can be defined once for the whole surface (for example by a choice of four surface points), while the local differential measure suggested by Lappin and Craft (2000) requires a set of five local reference points for each surface point under consideration. Second, eventual reconstruction of the surface shape is trivial in the 'referential' description, while for any 'differential' representation it necessarily involves a complicated integration process.

There is another very important benefit of using disparity defined with respect to a surface-centered reference frame. Given that projection of a surface in each eye is locally approximately orthographic (parallel), such a description can be used to construct a 3D surface representation that has invariant properties under rotations of the surface or movements of the observer relative to the surface.

This representation can be formed in what we call a 'cyclopean space' here. The cyclopean space is observer-centered, and a real-world point is represented in it by its two-dimensional cyclopean projection (i.e. its retinal position averaged between two eyes) plus the third coordinate, which is defined as its disparity (i.e. the difference between the two retinal projections). This is illustrated in Fig. 9 by the gaze-centered reference frame xyd. Basis vectors $\vec{x}$ and $\vec{y}$ are both orthogonal to the gaze direction while basis vector $\vec{d}$ is collinear with it.

Despite the actual retinal image being formed by central projection, parallel projection is a good approximation if the field of view is small. It is crucial to notice that when the retinal projections are approximately parallel, real-world Euclidean space is mapped into the cyclopean space by a linear transformation.[3] The changes that occur when the observer moves or changes his or her gaze can all be described by this linear transformation. In particular, it means that the coordinates of all surface points are transformed in exactly the same way.

Given a surface defined by two retinal images it is easy to construct a surface-based reference frame, as was described by Koenderink & van Doorn (1991) in their structure-from-motion algorithm (the two images in their

[3] The transformation is linear because parallel projection to both retinas and subsequent summation (for x and y components) and subtraction (for disparity component d) of the two retinal images are linear operations.

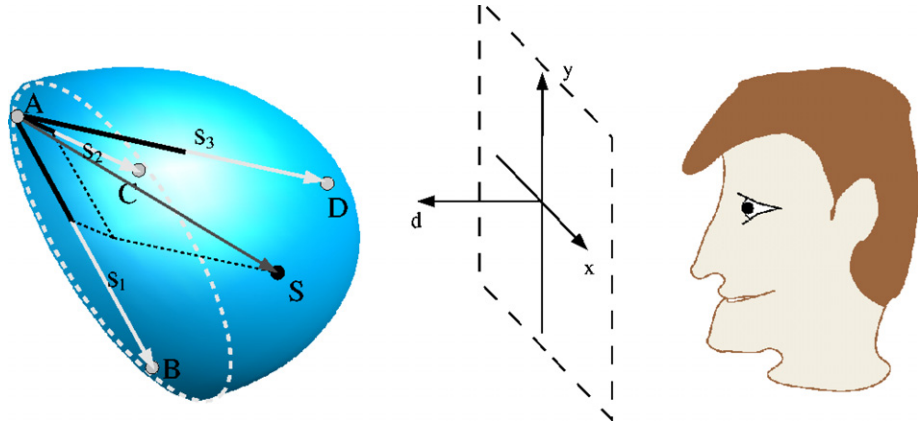

Fig. 9. A surface-centered reference frame defined by four noncoplanar surface points *A*, *B*, *C*, and *D*. Coordinates of a surface point, *S*, in this reference frame are given by parallel projections along the corresponding basis vectors as indicated by dashed lines. The text describes how this reference frame could be computed given the cyclopean direction (*xy*) and disparity (*d*) of each point and how the surface based coordinates are invariant to movement of the observer with respect to the surface.

paper were consecutive motion frames). Any four non-coplanar surface points *A*, *B*, *C*, and *D* define three basis vectors $\vec{s}_1$, $\vec{s}_2$, and $\vec{s}_3$ (Fig. 9). Point *A* serves as an origin, while the others form three basis vectors. Coordinates of a given surface point, *S*, are obtained by parallel projections along the basis vectors as shown by dashed lines in Fig. 9 . The resulting coordinates are equal to the length of the black bars drawn along each basis vector.

The advantage of the surface-centered frame is that surface rotations and viewpoint transformations affect the points defining the reference frame in exactly the same way as all the other surface points. Therefore, described in terms of these basis vectors, the surface representation is invariant to viewing transformations.

We have described the surface basis vectors, $\vec{s}_1$, $\vec{s}_2$, and $\vec{s}_3$, as 3D vectors in 'cyclopean space', *xyd*. However, in theory, the basis vectors could equally be derived from measures of the relative positions of reference points made in each monocular image, just as Koenderink and van Doorn described for a two-frame sequence (see also Glennerster & McKee, 2004). In either case, as an observer moves round an object, the surface-centered coordinates of point *S* using the basis vectors $\vec{s}_1$, $\vec{s}_2$, and $\vec{s}_3$ remain the same after the viewing transformation.

The invariance of this type of bas-relief, surface-based representation to an object's rotations and viewing transformations holds despite the fact that it is not a full, metric reconstruction. In many tasks a bas-relief representation is adequate (Glennerster, Rogers, & Bradshaw, 1996; Koenderink & van Doorn, 1991; Petrov & Glennerster, 2004; Tittle, Todd, Perotti, & Norman, 1995). Importantly, recovery of metric structure was not required in our experiments, and, accordingly, a bas-relief, surface-based representation was adequate to explain performance.

4.4. Conclusion

Stereoscopic detection of depth relief was studied using a 2D depth profile (defined by a triangle composed of one target dot and two reference dots), and a surface depth 'bump' (defined by a tetrahedron composed of one target dot and three reference dots). Based on our previous work, two disparity cues were proposed as main factors limiting performance of this task: (a) disparity gradient difference and (b) disparity relative to the reference line (reference plane for the surface stimulus). Performance for both types of depth relief was well described by the latter cue, but was found to be inconsistent with the predictions based on the disparity gradient difference cue. Although there were some circumstances in which the relative disparity or disparity gradient between points was an important factor (when target and reference were close to one another), we found that in general it is the disparity of a point with respect to a line or plane that determines stereoacuity performance.

Acknowledgments

We thank Andrew Parker and Suzanne McKee for helpful discussions. This work was supported by the Welcome Trust (Grant 056657).

Appendix A

This section gives formulae defining the two disparity cues discussed in the paper: disparity gradient difference ω between the target and reference dots *A* and *B* and disparity, *s*, with respect to the reference line *AB*.

Referring to Fig. 10, the two disparity gradients between the reference dots *A* and *B* and the target dot *T* are given by

$$\nabla^l = 2\frac{\Delta+\delta}{D-d} \quad \text{and} \quad \nabla^s = 2\frac{\Delta-\delta}{D+d}, \tag{1}$$

where ∇^l and ∇^s stand for the larger and smaller of the two gradients, respectively. For their difference:

$$\omega \equiv \nabla^l - \nabla^s = 4\,\frac{\delta D + \Delta d}{D^2 - d^2}. \tag{2}$$

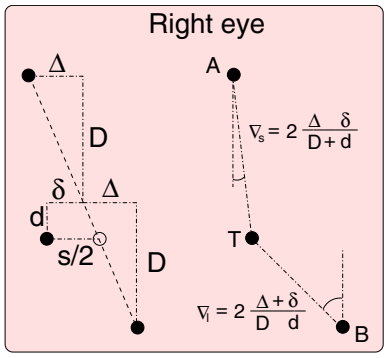

Fig. 10. The geometry of the 2D stimulus for the right eye. The left eye stimulus is a mirror reflection of the right eye stimulus. Δ, D, d, and δ indicate the relative positions of the dots in monocular images. ∇^s (small) and ∇_l (large) stand for two disparity gradients between the target T and the reference dots A and B, respectively.

It is easy to see that the target disparity s with respect to the line AB is given by: $s/2 = \delta + d\Delta/D$, and Eq. (2) can be rewritten as

$$\omega = s\frac{2D}{D^2 - d^2} = \frac{2s}{D}\frac{1}{1-x^2}, \tag{3}$$

where the relative displacement d/D was denoted by x in the last equality. This shows that ω is proportional to s, and, given a constant s, the disparity gradient difference ω depends on the target displacement along line AB as $1/(1-x^2)$. This functional form was multiplied by a free parameter to fit data in Experiments 1, 2 and 3.

References

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology—London, 211*, 599–622.

Glennerster, A., & McKee, S. P. (1999). Bias and sensitivity of stereo judgements in the presenece of a slanted reference plane. *Vision Research, 39*, 3057–3069.

Glennerster, A., & McKee, S. P. (2004). Sensitivity to depth relief on slanted surfaces. *Journal of Vision, 4*, 378–387.

Glennerster, A., McKee, S. P., & Birch, M. D. (2002). Evidence of surface-based processing of binocular disparity. *Current Biology, 12*, 825–828.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research, 36*, 3441–3456.

Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research, 43*, 2539–2558.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of Optical Society of America A—Optics Image Science and Vision, 8*, 377–385.

Lappin, J. S., & Craft, W. D. (2000). Foundations of spatial vision: from retinal images to perceived shapes. *Psychological Review, 107*, 6–38.

McKee, S. P., Welch, L., Taylor, D. G., & Bowne, S. F. (1990). Finding the common bond: stereoacuity and the other hyperacuities. *Vision Research, 30*, 879–891.

Petrov, Y., & Glennerster, A. (2004). The role of the local reference in stereoscopic detection of depth relief. *Vision Research, 44*, 367–376.

Rogers, B. J., & Cagenello, R. (1989). Disparity curvature and the perception of three-dimensional surfaces. *Nature, 339*, 135–137.

Rogers, B. J., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22*, 261–270.

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). A hierarchical analysis of alternative representations in the perception of 3-D structure from motion and stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 663–678.

Tyler, C. W. (1974). Depth perception in disparity gratings. *Nature, 251*, 140–142.

Westheimer, G., & McKee, S. P. (1979). What prior uniocular processing is necessary for stereopsis? *Investigative Ophtalmology and Visual Science, 18*, 614–621.